❒ 2922

Nutrition information estimation from food photos using machine learning based on multiple datasets

Mustafa Al-Saffar, Wadhah Baiee
Department of Software, Information Technology, University of Babylon, Babylon, Iraq

Article Info	ABSTRACT
	Bodyweight, blood pressure, and cholesterol are all risk variables that can aid people in making educated decisions regarding their health promotion activities. Food choices are among the most effective methods for preventing chronic illnesses, including heart disease, diabetes, stroke, and some malignancies. Because various meals give varying amounts of energy and minerals, good eating necessitates keeping track of the nutrients we ingest. Furthermore, there is a paucity of information on whether understanding food constituents might aid in more accurate nutrition calculations. Therefore, this research suggests processing food images on social media to anticipate the contents of each food and extracting nutrition information for each food image to serve as healthy implicit feedback to take advantage of the rapid accumulation of rich photos on social media. The proposed methodology is a framework based on a machine-learning model for predicting food ingredients. We also compute critical health metrics for each ingredient and combine them to obtain nutrition data for the food. The result revealed a promising way of extracting food components and nutrition information. Compared with other researchs, our proposed prediction and attribute extraction strategy achieves a remarkable accuracy of 85%. *This is an open access article under the [CC BY-SA](#) license.*

Corresponding Author:

Mustafa Al-Saffar
Department of Software, Information Technology, University of Babylon
80-street, Hilla, Babil, Iraq
Email: mustafaalsaffar.sw.msc@student.uobabylon.edu.iq

1. INTRODUCTION

Food significantly affects humans' quality of life [1], health, and happiness [2]. The number of overweight or obese persons is increasing. According to the WHO [3], over 1.9 billion obese adults between the ages of 18 and over 650 million obese adults, obesity is a significant cause of diseases. For these reasons, food-related research [4]–[6] has consistently been a hot topic and garnered significant attention from various sectors. Previously, food-related research focused on various topics, including food selection [7] and food perception [8]. These studies, however, were undertaken before the web transformed research in various fields.

Additionally, most approaches rely on small data sets, such as questionnaires, cookbooks, and recipes [9]. Nowadays, with the rapid rise of multiple networks such as social networks, mobile networks, and the Internet of Things, people can easily share food images, recipes, cooking videos, and meal diaries, resulting in vast food databases [10]. These food data suggest a wealth of knowledge and hence present significant prospects for food-related research, including the discovery of food perception principles [11], the analysis of culinary habits [6], and diet monitoring [5]. Additionally, network analysis, computer vision, machine learning, and data mining offer various unique data analysis tools. Recent breakthroughs in artificial intelligence (AI), notably in deep learning [8], have sparked renewed interest in large-scale food-related

studies [9], [10], owing to AI's improved ability to learn representations from several different sorts of data [12]. The research problems for this study boil down to two main points; the first is to extract nutrition information for each food ingredient from food photos. The second is that most past studies have not merged the computed important health metrics for each ingredient in the food meal photo to provide nutrition data for the food, which is a nice balance that our research will achieve. Most research on predicting nutritional value from food images has focused on taking an input food image and classifying it using a trained list of food categories via convolutional neural networks (CNN) layers [13], [14]. Then, based on the identified category, it displays the food item's projected nutritional value. Many researchers are interested in solving image food prediction problems and extracting nutritional value predictions from them [15]. Researchers have only lately begun to work on predicting food images and extracting predicting nutritional data. The following are the most often conducted studies in this field:

− Deepak *et al.* [16] developed a model that takes an input food image and classifies it using CNN layers according to the training list of food categories. Then, based on the selected category, it displays the estimated nutritional value of the food item. In this scenario, they recognized the food image using CNN, a deep learning approach, bypassing it through layers such as Dense, Dropout, Flatten, Conv2D, and Maxpooling2D. Additionally, they designed a system that uses an ever-growing and dynamic collection of culinary images. They know that food and its classes are immense and constantly growing, resulting in the disastrous loss of idealistic notions in present systems. To my mind, this article introduces a powerful photo prediction technique that may be used to identify the sort of food without relying on meal ingredients.

− Hong *et al.* [17] used deep learning neural networks to create a novel approach for automatically predicting meal calories from ingredient pictures. The method begins by training an object recognition model to recognize all the food elements in the image, using the white dish as a reference object. It then uses polynomial linear regression to fit the relationship between the weight of the food ingredient and the area of the food ingredient in the image. Finally, it estimates the calories from the ingredients' calorific values. The article's shortcoming is that it estimates food calories using constituent photographs rather than meal shots.

− Amiri *et al.* [18] created methods for multiple regression, including least squares linear regression (linear) with l2 regularization (ridge). Additionally, the state-of-the-art approach for deriving nutrition facts from food descriptions as described in–a CNN employed word n-grams to match food items to the USDA dataset to generate nutrition facts. They expanded on this strategy by teaching each other the nutrition data and ingredients of various foods. They built a framework for multi-task learning to enable concurrent learning of ingredients and nutrition data from meal descriptions. To my mind, this paper is that it estimates food calories using meal descriptions rather than meal photos (text analysis).

− In [19] and [20] proposed that SVM and enhanced MLP models are used in this research to perform food item recognition and calorie prediction. The suggested study utilizes a variety of preprocessing approaches, segmentation, and feature extraction to analyze a single food item. For recognition, the collected features are input into SVM and MLP classifiers. In my opinion, the paper's weak point is that it estimates food calories based on raw shots of food, not meal photographs. However, most of these studies have not considered the computed critical health metrics in the food meal photo.

The proposed methodology is into two distinct ways of processing, and the first is the use of learning-based CNN models for ingredient prediction. In contrast, the second generates a list of predicted ingredients and nutrition facts estimation.

This study proposes a vision of food photos for predicting food ingredients. The researchers also compute critical health metrics for each ingredient and combine them to obtain nutrition data for the food. This study proposes a visual representation of food images to predict meal constituents. Additionally, the researchers compute essential health metrics for each ingredient and combine them to obtain the food's nutritional information.

2. THE PROPOSED APPROACH

Our model's primary objective is to develop a framework based on the machine-learning framework for predicting food ingredients. Additionally, the researchers computed essential health metrics for each ingredient and combined them to obtain the food's nutritional information. The findings demonstrated a promising method for collecting food components and nutrition information from them, allowing developers and architects to leverage this model when designing food and health systems and systems of recommendation.

Figure 1 illustrates an example of the proposed model. The model has the main module of the CNN-based pre-trained model. After the predicting procedure is complete using a pre-trained model, the estimated

nutrition information model gives the item's name and constituents. Nutritional data will be derived from the Nutrition5k dataset. Table 1 shows result of inverse cooking model. Table 2 shows result of nutrition information estimation.
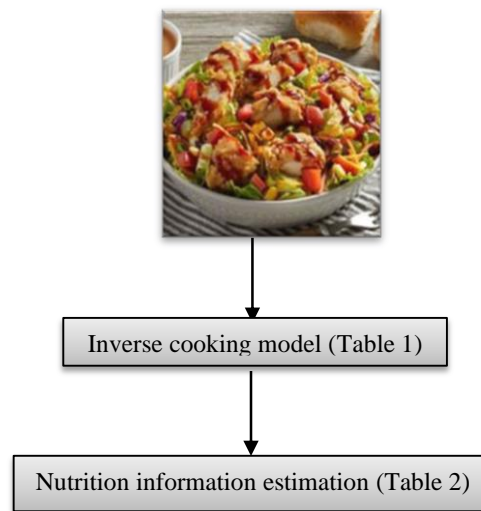

Figure 1. Proposed model example

Table 1. Result of inverse cooking model

Title	Ingredients
Bbq chicken salad	Onion, chicken, pepper, lettuce, barbecue_sauce

Table 2. Result of nutrition information estimation

ingr	Cal/g	Fat/g	Carb/g	Protein/g
Onion	0.4	0.001	0.09	0.011
Chicken	1.65	0.036	0	0.31
Pepper	0.4	0.002	0.093	0.02
Lettuce	0.15	0.002	0.029	0.014
Barbecue_sauce	1.72	0.006	0.41	0.008
Total	4.32	0.047	0.622	0.363

3. METHOD
3.1. Datasets
3.1.1. Yelp dataset

The primary resource in this model is the Yelp dataset, which will estimate nutrition facts in this application. A subset of Yelp's businesses, reviews, and user data is available for personal, educational, and research use. Available in the form of JSON files. It contains a directory of 10,000 No restaurants, complete with addresses, menu selections, and ratings. Yelp provides around 280,000 images from over 2000 establishments, with the dataset available on the Yelp data challenge website. This dataset contains interior, exterior, beverage, and food photographs. Yelp has designated the four categories above, but there are no subcategories for specific types of cuisine. Customers or business owners upload the majority of photographs. These photographs may contain good captions which accurately describe the photographs they accompany. Numerous photographs lack captions or contain inaccurate captions [21]. The researchers preprocess the dataset by extracting only food photos and using them to test the model.

3.1.2. Nutrition5k dataset

Nutrition5k is a visual and nutritional data dataset for ~5k realistic plates of food captured from Google cafeterias using a custom scanning rig. This study is releasing this dataset alongside our recent CVPR 2021 paper to help promote research in visual nutrition understanding [22]. The researchers used this dataset to estimate nutrition facts for each ingredient in the meal.

3.2. Model design and implementation

The proposed methodology is into two distinct ways of processing, and the first is the use of learning-based CNN models for ingredient prediction, as shown in Figure 2. The second generates a list of predicted ingredients and nutrition facts estimation, as shown in Figure 3.

3.2.1. Pre-trained convolutional neural network model

The model used consists of two main phases; the first phase is called Image Encoder; this phase is based on resnet50 CNN architecture to extract features from food photos, and the second phase is called ingredient decoder; this phase is based on the transformer model to generate a list of ingredients as shown in Figure 2.
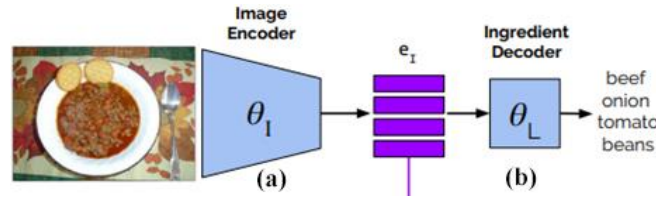

Figure 2. Trained model, (a) image encoder using ResNet50 and (b) ingredient decoder using transformal model [23]

A *list of ingredients* is a varying in size, a systematic gathering of distinct meal ingredients. Training data consists of M image and ingredient list pairs $\{(\mathbf{x}^{(i)}, \mathbf{L}^{(i)})\}_{i=0}^{M}$. the goal is to predict $\hat{\mathbf{L}}$ from an image $\mathbf{x}$ by maximizing the following objective [23]:

$$arg \max_{\theta_I, \theta_L} \sum_{i=0}^{M} log \ p\left(\hat{L}^{(i)} = L^{(i)} \mid x^{(i)}; \theta_I, \theta_L\right) \tag{1}$$

A *set of ingredients* is a variable-sized, disorganized grouping of distinct meal ingredients. The model can obtain a set of ingredients S by selecting K ingredients from the dictionary $\mathcal{D}: S = \{s_i\}_{i=0}^{K}$. Training data consists of M image and ingredient set pairs: $\{(\mathbf{x}^{(i)}, \mathbf{s}^{(i)})\}_{i=0}^{M}$. In this case, the goal is to predict $\hat{\mathbf{s}}$ from an image $\mathbf{x}$ by maximizing the following objective [23]:

$$arg \max_{\theta_I, \theta_L} \sum_{i=0}^{M} log \ p\left(\hat{s}^{(i)} = s^{(i)} \mid x^{(i)}; \theta_I, \theta_L\right) \tag{2}$$

The researchers employed a target distribution technique $p(\mathbf{s}^{(i)} \mid \mathbf{x}^{(i)}) = \mathbf{s}^{(i)}/\sum_j \mathbf{s}_j^{(i)}$ to model the joint distribution of set elements and train a model by minimizing the cross-entropy loss between $p(\mathbf{s}^{(i)} \mid \mathbf{x}^{(i)})$ and the model's output distribution $p(\hat{\mathbf{s}}^{(i)} \mid \mathbf{x}^{(i)})$. Nonetheless, it is unclear how to transform the target distribution back to the set of items with variable cardinality corresponding to it. In this case, The researchers built a feed-forward network and trained it with the target distribution cross-entropy loss. To recover the ingredient set, The researchers proposed to sample elements probabilities greedily $p(\hat{\mathbf{s}}^{(i)} \mid \mathbf{x}^{(i)})$ and stop the sampling once the sum of probabilities $p(\hat{\mathbf{s}}^{(i)} \mid \mathbf{x}^{(i)})$ Furthermore, stop the sampling once the sum of probabilities of selected elements is above a threshold. The researchers referred to this model as *feed-forward (target distribution)* [23].

This paper used A model that has been pre-trained using feed-forward convolutional networks to predict sets of ingredients (as shown above). Several losses have been used to train these models, including binary cross-entropy, soft intersection over union, and target distribution cross-entropy. Notably, binary cross-entropy is the only one that ignores the set's dependencies. On the other hand, sequential models predict lists, impose order, and take advantage of element dependencies. Finally, this study explores newly proposed models that combine set and cardinality prediction to choose which components to include in the set.

3.2.2. Nutrition information estimation

Figure 3 illustrates that before predicting our dataset's pictures, it must be PreProcess Stage to the dataset to extract food photos only from the Yelp dataset. The next stage is predicted food ingredients. In this

stage, the researchers called the pretrained model to predict food ingredients from food photos. In the final stage, the nutritional information was estimated stage. After obtaining the name of the food and its components, it was sent to the in estimating the nutritional components. After taking the predicted components, the nutritional information for each component was calculated from the Nutrition5k dataset. All the ingredients were collected together to extract the essential nutritional information (calories, carbohydrates, fat, protein), according to the United States Department of Agriculture Food and Nutrition Information Center [24] for one gram of the food, as shown in Algorithm 1. The mass and size of the food were not considered due to the difficulty of predicting the size of the food and its components from the picture.
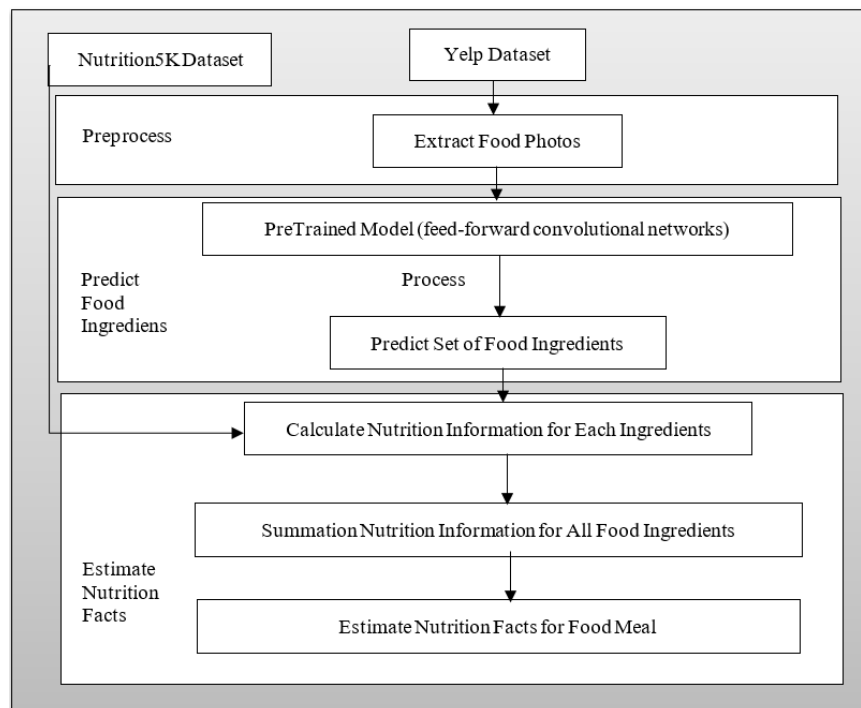

Figure 3. Block diagram for the proposed model

Algorithm 1 : Nutrition Information Estimation :
Inputs : photo, photo id, photo category, ingredient name, calories, fat, carbohydrate ,protein
                                                                                        // from Dataset

Outputs : result cal , result fat , result carb ,result prot
Process :
Begin
    FOREACH photo category is "food" DO
          ResultIngedientsList ← PredictFoodIngedients(photo)
    FOREACH ResultIngedient in ResultIngedientsList DO
        IF ResultIngedient within IngedientsList(DS) THEN
            Summation result Cal with Calories
            Summation result fat with Fat
            Summation result carb with Carbohydrate
            Summation result prot with Protein
        ELSE
            Summation result Cal with zero
            Summation result fat with zero
            Summation result carb with zero
            Summation result prot with zero
End

---

4. RESULTS AND DISCUSSION

Table 3 illustrates that the model's test accuracy examples depend on the correct prediction of components. The more correct the estimate of the components, the more accurate and better the result of calculating nutritional information. The researchers assess the quality of anticipated elements using user trials to evaluate the model's effectiveness. They compare their model's performance to that of humans in ingredient production by randomly selecting 15 images from the test set and asking participants to choose up to 21 different ingredients that correspond to the presented image. To lessen the work's complexity for humans, they reduced the vocabulary for ingredients from 1489 to 220 by increasing the frequency threshold from ten to one thousand. They received responses from 31 unique individuals, averaging 5.6 responses per image. They retrain our top ingredient prediction algorithm on the restricted vocabulary of components to ensure a fair comparison. They compute IoU, and F1 ingredient evaluations gathered from humans, the retrieval baseline, and their methodology. The findings emphasize the complexity of the endeavor. Humans outperform the retrieval reference standard (35.25% F1 vs. 30.60%). (F1 of 35.25% vs. 30.6%, respectively) [25]. Additionally, their technique beats baseline human performance and retrieval-based systems, with an F1 score of 49.09%. The extra material contains qualitative comparisons of synthetic and human-written ingredients (including ingredients from ordinary and skilled users).

Table 3. Ingredient prediction examples

Meal food photo	Ingredient's production	Ingredients real	Estimated nutritional information/g	Real nutritional information/g
	Onion, chicken, pepper, lettuce, barbecue_sauce	Chicken, tomatoes, onion, perpper, lettuce, beans, corn, BBQ sauce	Cal/g=4.32 Fat/g=0.047 Carb/g=0.62 Prot/g=0.36	Cal/g=7.14 Fat/g=0.11 Carb/g=1.11 Prot/g=0.46
	Oil, pepper, onion, seeds, spinach, vinegar, salt, egg, sugar	Spinach, tahini, miso pasta, sesame, oil, caster suger, lime juice, toasted sesame seeds	Cal/g=20.91 Fat/g=1.604 Carb/g=1.33 Prot/g=0.49	Cal/g=28.48 Fat/g=2.089 Carb/g=2.09 Prot/g=0.68

Figure 4 shows the high predictive accuracy rate of the nutritional estimation model of the food images after taking 53 samples, inserting them into the proposed model, and comparing the estimated ingredients from the proposed model with the actual ingredients of the food. Worthy of attention is that the accuracy of the proposed model depends mainly on the accuracy of the ingredient's prediction model.
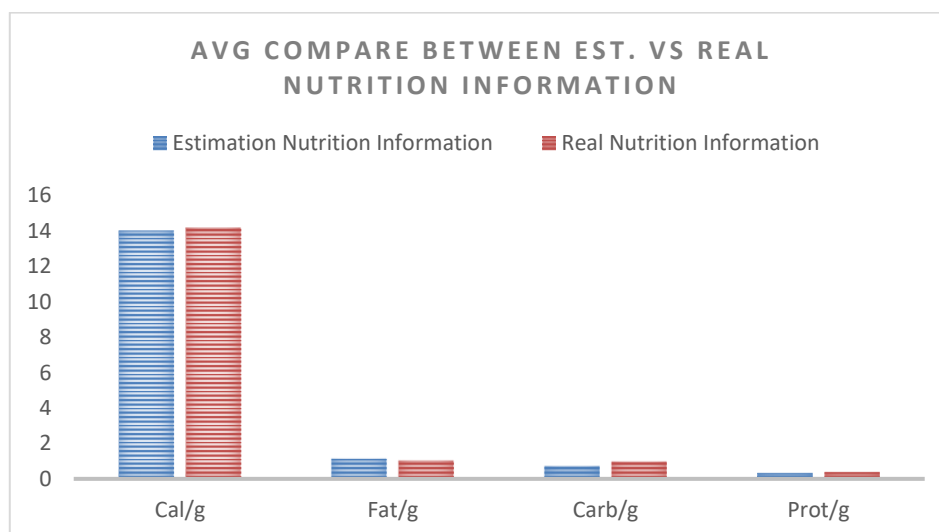

Figure 4. Accuracy rate ingredient prediction for proposed model

5.   CONCLUSION

At the moment, obesity is a severe worry of human existence, and people have developed an interest in monitoring their weight and eating habits to avoid becoming obese. Thus, this research presents a revolutionary approach for informing us about the type of food we consume and its characteristics. This research developed a system for nutritional assessment that takes an image of a meal and generates a title and ingredients list. This is the first study to estimate food ingredients from food photographs and evaluate nutritional information based on the predicted ingredients. The model will inform us of the dish's characteristics. Our model was built using a dataset from Yelp and Nutrition5k that contains an average meal. This study includes a predicting model for food items and a method for quantifying the food's attributes through an attribute estimation model. Our proposed prediction and attribute extraction strategy achieves a remarkable accuracy of 85%. Additionally, this research discussed prospective modifications and future work to improve the model's utility and accuracy by increasing component prediction accuracy and considering food mass when estimating nutrition information.

REFERENCES

[1]   J. Li, R. Guerrero, and V. Pavlovic, "Deep cooking: Predicting relative food ingredient amounts from images," *MADiMa 2019 - Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management, co-located with MM 2019*. pp. 2–6, 2019. doi: 10.1145/3347448.3357164.
[2]   P. Achananuparp, E. P. Lim, and V. Abhishek, "Does journaling encourage healthier choices? Analyzing healthy eating behaviors of food journalers," *ACM International Conference Proceeding Series*, vol. 2018-April. pp. 35–44, 2018. doi: 10.1145/3194658.3194663.
[3]   "WHO | World Health Organization." https://www.who.int/en (accessed May 10, 2022).
[4]   L. Canetti, E. Bachar, and E. M. Berry, "Food and emotion," *Behavioural Processes*, vol. 60, no. 2. pp. 157–164, 2002. doi: 10.1016/S0376-6357(02)00082-7.
[5]   J. Chung, J. Chung, W. Oh, Y. Yoo, W. G. Lee, and H. Bang, "A glasses-type wearable device for monitoring the patterns of food intake and facial activity," *Scientific Reports*, vol. 7. 2017. doi: 10.1038/srep41690.
[6]   S. Sajadmanesh *et al.*, "Kissing cuisines: Exploring worldwide culinary habits on the web," *26th International World Wide Web Conference 2017, WWW 2017 Companion*. pp. 1013–1021, 2017. doi: 10.1145/3041021.3055137.
[7]   M. Nestle *et al.*, "Behavioral and social influences on food choice," *Nutrition Reviews*, vol. 56, no. 5, pp. 50–64, 1998. doi: 10.1111/j.1753-4887.1998.tb01732.x.
[8]   L. B. Sørensen, P. Møller, A. Flint, M. Martens, and A. Raben, "Effect of sensory perception of foods on appetite and food intake: A review of studies on humans," *International Journal of Obesity*, vol. 27, no. 10. pp. 1152–1166, 2003. doi: 10.1038/sj.ijo.0802391.
[9]   S. Phiphiphatphaisit and O. Surinta, "Food Image Classification with Improved MobileNet Architecture and Data Augmentation," *Proceedings of the 2020 The 3rd International Conference on Information Science and System*, Cambridge, United Kingdom, 2020, pp. 51-56, doi: 10.1145/3388176.3388179.
[10]   M. A. Subhi, S. H. Ali, and M. A. Mohammed, "Vision-Based Approaches for Automatic Food Recognition and Dietary Assessment: A Survey," *IEEE Access*, vol. 7. pp. 35370–35381, 2019. doi: 10.1109/ACCESS.2019.2904519.
[11]   O. G. Mouritsen, R. Edwards-Stuart, Y. Y. Ahn, and S. E. Ahnert, "Data-driven methods for the study of food perception, preparation, consumption, and culture," *Frontiers in ICT*, vol. 4, May 2017. doi: 10.3389/fict.2017.00015.
[12]   L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, "Application of Deep Learning in Food: A Review," *Comprehensive Reviews in Food Science and Food Safety*, vol. 18, no. 6, pp. 1793-1811, 2019, doi: 10.1111/1541-4337.12492.
[13]   R. Yunus *et al.*, "A Framework to Estimate the Nutritional Value of Food in Real Time Using Deep Learning Techniques," *IEEE Access*, vol. 7. pp. 2643–2652, 2019. doi: 10.1109/ACCESS.2018.2879117.
[14]   S. F. Situju, H. Takimoto, S. Sato, H. Yamauchi, A.Kanagawa, and A. Lawi, "Food Constituent Estimation for Lifestyle Disease Prevention by Multi-Task CNN," *Applied Artificial Intelligence*, vol. 33, no. 8, pp. 732–746, 2019. doi: 10.1080/08839514.2019.1602318.
[15]   V. H. Reddy, S. Kumari, V. Muralidharan, K. Gigoo, and B. S. Thakare, "Literature Survey—Food Recognition and Calorie Measurement Using Image Processing and Machine Learning Techniques," *Lecture Notes in Electrical Engineering*, vol. 570. pp. 23–37, 2020. doi: 10.1007/978-981-13-8715-9_4.
[16]   N. R. Deepak, G. K. Suhas, B. Bhagappa, and P. K. Pareek, "A Framework for Food recognition and predicting its Nutritional value through Convolution neural network," *SSRN Electronic Journal*. 2022. doi: 10.2139/ssrn.4040968.
[17]   H. Liang, Y. Gao, Y. Sun, and X. Sun, "CEP- calories estimation from food photos," *International Journal of Computers and Applications*, vol. 42, no. 6, pp. 569-577, 2018. doi: 10.1080/1206212X.2018.1486558.
[18]   H. Amiri, A. L. Beam, and I. S. Kohane, "Learning to Estimate Nutrition Facts from Food Descriptions," AMIA, pp. 1–2, 2019.
[19]   T. V. Yadalam, V. M. Gowda, V. S. Kumar, D. Girish and N. M., "Career Recommendation Systems using Content based Filtering," *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, pp. 660-665, doi: 10.1109/ICCES48766.2020.9137992.
[20]   R. D. Kumar, E. G. Julie, Y. H. Robinson, S. Vimal, and S. Seo, "Recognition of food type and calorie estimation using neural network," *Journal of Supercomputing*, vol. 77, no. 8. pp. 8172–8193, 2021. doi: 10.1007/s11227-021-03622-w.
[21]   "Yelp Dataset," https://www.yelp.com/dataset (accessed May 14, 2022).
[22]   "google-research-datasets/Nutrition5k: Detailed visual + nutritional data for over 5,000 plates of food." https://github.com/google-research-datasets/Nutrition5k (accessed May 14, 2022).
[23]   A. Salvador, M. Drozdzal, X. Giro-i-Nieto, and A. Romero, "Inverse Cooking: Recipe Generation from Food Images," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10453-10462, 2019.
[24]   "USDA." https://www.usda.gov/ (accessed May 20, 2022).
[25]   Zhao-Yan Ming, J. Chen, Y. Cao, C. Forde, Chong-Wah Ngo, and T. S. Chua, "Food Photo Recognition for Dietary Tracking: System and Experiment," *International Conference on Multimedia Modeling,* Springer, Cham, pp. 129-141, 2018, doi: 10.1007/978-3-319-73600-6_12.

BIOGRAPHIES OF AUTHORS

Mustafa Al-Saffar ⒾⒹ 🅖 ⓈⒸ Ⓟ is a Teacher in the Ministry of Education, Iraq. He undertook his BSc in Software Department at the University of Babylon, Iraq. Currently, He is studying MSc in the Software Department. His area of research focuses on information systems in recommender systems. He can be contacted at email: mustafaalsaffar.sw.msc@student.uobabylon.edu.iq.

Wadhah R. Baiee ⒾⒹ 🅖 ⓈⒸ Ⓟ is a lecturer in the College of Information Technology at the University of Babylon, Iraq. He undertook his Ph.D. in Computer Science at the University of Babylon, Iraq. His area of research focuses on information systems and their applications in recommender systems and GIS. He can be contacted at email: wadhah.baiee@uobabylon.edu.iq.